Title: Burden and profile of somatic mutation in duodenal adenomas from patients with familial adenomatous- and MUTYH-associated polyposis

Laura E. Thomas[1], Joanna J. Hurley[1,2], Elena Meuser[1], Sian Jose[1], Kevin E. Ashelford[1], Matthew Mort[1], Shelley Idziaszczyk[1], Julie Maynard[1], Helena Leon Brito[1], Manon Harry[1], Angharad Walters[1], Meera Raja[1], Sarah-Jane Walton[3], Sunil Dolwani[1,4], Geraint T. Williams[1], Meleri Morgan[5], Morgan Moorghen[3,7], Susan K. Clark[3,6], Julian R. Sampson[1]

1. Institute of Medical Genetics, Division of Cancer and Genetics, Cardiff University, School of Medicine, Cardiff, CF14 4XN, UK
2. Department of Gastroenterology, Prince Charles Hospital, Merthyr Tydfil, CF47 9TD, UK.
3. The Polyposis Registry, St Marks Hospital, Watford Road, Harrow HA1 3UJ, UK
4. Division of Population Medicine, Cardiff University School of Medicine, Cardiff CF14 4XN, UK
5. Department of Pathology, University Hospital for Wales, Cardiff, CF14 4XN, UK
6. Department of Surgery and Cancer, Faculty of Medicine, Imperial College, London SW7 2AX, UK
7. Department of Pathology, St Marks Hospital, Watford Road, Harrow HA1 3UJ UK

Running title: Somatic mutational profile in FAP and MAP duodenal adenomas

Keywords: FAP; MAP; Adenoma; Exome; Duodenum

Financial support:

This project has been funded by the Welsh Government through Health and Care Research Wales through a NISCHR Fellowship to LET and by the Wales Gene Park and Wales Cancer Research Centre.

Correspondence to:

Professor Julian Sampson, Institute of Medical Genetics, Cardiff University School of Medicine, Heath Park Way, Cardiff, CF14 4XN, UK. Phone: +44 2920 748170. Fax: +44 2920745661. Mail: Sampson@cf.ac.uk

Conflicts of interest:

The authors declare no potential conflicts of interest.

Word count: 4185

References: 47

Total number of tables and figures: 5

Supplementary files: 7

Statement of translational relevance:

Surveillance duodenoscopy is undertaken in patients with familial adenomatous polyposis (FAP) or MUTYH-associated polyposis (MAP) to reduce the risk of duodenal cancer. Current guidelines in the USA and Europe recommend that the screening interval and decisions on interventions are based upon Spigelman staging of duodenal polyposis. In this study we demonstrate a greater mutational burden in MAP than FAP duodenal adenomas despite lower Spigelman stage duodenal polyposis in the MAP patients studied. These findings suggest that the risk of progression to cancer in the context of early stage duodenal polyposis could be higher in MAP than FAP patients and challenge the assumption that the same surveillance protocols should be applied in MAP and FAP.

Abstract

Purpose: Duodenal polyposis and cancer are important causes of morbidity and mortality in familial adenomatous polyposis (FAP) and MUTYH-associated polyposis (MAP). This study aimed to comprehensively characterize somatic genetic changes in FAP and MAP duodenal adenomas to better understand duodenal tumorigenesis in these disorders.

Experimental Design: Sixty-nine adenomas were biopsied during endoscopy in 16 FAP and 10 MAP patients with duodenal polyposis. Ten FAP and 10 MAP adenomas and matched blood DNA samples were exome sequenced,

42 further adenomas underwent targeted sequencing and 47 were studied by array comparative genomic hybridization. Findings in FAP and MAP duodenal adenomas were compared to each other and to the reported mutational landscape in FAP and MAP colorectal adenomas.

Results: MAP duodenal adenomas had significantly more protein-changing somatic mutations (P = 0.018), truncating mutations (P = 0.006) and copy number variants (P = 0.005) than FAP duodenal adenomas, even though MAP patients had lower Spigelman stage duodenal polyposis. Fifteen genes were significantly recurrently mutated. Targeted sequencing of *APC*, *KRAS*, *PTCHD2* and *PLCL1* identified further mutations in each of these genes in additional duodenal adenomas. In contrast to MAP and FAP colorectal adenomas, neither exome nor targeted sequencing identified WTX mutations (P=0.0017).

Conclusions: The mutational landscapes in FAP and MAP duodenal adenomas overlapped with, but had significant differences to those reported in colorectal adenomas. The significantly higher burden of somatic mutations in MAP than FAP duodenal adenomas despite lower Spigelman stage disease could increase cancer risk in the context of apparently less severe benign disease.

Introduction

Familial adenomatous polyposis (FAP) and *MUTYH*-associated polyposis (MAP) are inherited disorders characterized by colorectal polyposis and cancer. They are also associated with extra-colonic manifestations including polyposis and cancer in the upper gastrointestinal tract, most notably duodenal disease that has become an important cause of morbidity and mortality as the management of colorectal disease has improved (1). A recent study of FAP estimated the lifetime risk of duodenal polyposis to be 88% and of cancer to be 18% (2). In a multicenter retrospective study of MAP, duodenal polyps were noted in 26 of 150 (17%) patients undergoing duodenoscopy and the lifetime risk of duodenal cancer was estimated at 4% (3). A more recent study in two specialist centers identified duodenal adenomas in 31 of 92 (34%) MAP patients undergoing endoscopy at a median age of 50 years (4).

In patients with FAP or MAP regular endoscopic surveillance of the duodenum has been advocated from the age of 25-30 years (1). Spigelman staging based upon the number, size, dysplasia and presence of villous histology of adenomas was developed to better define the severity of duodenal disease in FAP (5) and is recommended to guide the frequency of surveillance, stratify cancer risk and inform decisions about surgical intervention (6). Duodenal disease in FAP appears to progress slowly through Spigelman stages (0-IV) with an associated increase in cancer risk (7). The natural history of duodenal polyposis in MAP is not well defined but there are reports of duodenal cancer

occurring in the context of minimal background polyposis (3,8). More evidence is required to support or refute current recommendations to apply the same Spigelman stage-based surveillance and intervention for MAP as FAP (1,6).

Rapid recurrence of duodenal adenomas has been reported following endoscopic polypectomy in patients with FAP (9,10). Surgical treatments including ampullectomy, duodenectomy and pancreatico-duodenectomy appear effective for cancer prevention but are associated with significant procedure-associated risks (7,11). Medical treatment using the cyclooxygenase (COX) inhibitors sulindac and celecoxib has proven less effective in the duodenum than the colorectum (12–15) but a recent trail of combined COX and epidermal growth factor receptor inhibition with sulindac and erlotinib demonstrated promising short-term effects on duodenal polyp burden (16). The efficacy of medical and surgical treatment or prevention of duodenal disease in MAP remains unknown.

In colorectal tumorigenesis, the nature and positions of *APC* mutations appear to determine a critical level of over-activation of β-catenin signaling that leads to a failure in cell growth control without induction of apoptosis (17), a scenario described by the “just right” hypothesis (18). The situation in FAP-associated upper gastrointestinal tumors appears to be subtly different as somatic *APC* mutations cluster in a more 3’ region (19). Severe upper intestinal polyposis is also associated with a more 3’ location of inherited *APC* mutations (19).

Recently, a comprehensive survey of the mutational landscape of colorectal adenomas from patients with FAP and MAP was made using exome sequencing (20). This confirmed the importance of somatic *APC* and *KRAS* mutations as drivers of early colorectal tumorigenesis in both disease settings. It also identified frequent somatic mutations of *WTX* (also known as *FAM123B* and *AMER1*) as had been reported previously in sporadic colorectal cancer (21) and that, like *APC* mutations, may act through deregulation of β-catenin turnover. Although comprehensive molecular genetic studies of duodenal adenomas or carcinomas in patients with FAP have not been reported, targeted sequencing has confirmed a role for *APC* and revealed oncogenic mutations of *KRAS* in 9-30% of FAP duodenal adenomas (22–25). Comparable studies of MAP-associated duodenal tumors have not been reported.

In this study we applied whole exome and targeted Sanger sequencing and array comparative genomic hybridization (aCGH) to characterize somatic genetic variation associated with the development of duodenal adenomas in patients with FAP and MAP.

Materials and Methods

Patients and Samples

Ethical approval was granted by the UK NHS Research Ethics Committee system (reference 10/MRE093). All patients provided written informed

consent. This study was completed in accordance with the ethical guidelines of the Declaration of Helsinki. Their diagnoses of FAP or MAP were confirmed by genetic testing. Biopsies of approx. 3mm of duodenal polyps were taken during upper GI surveillance endoscopy. Spigelman stage was calculated using the method described by Saurin et al (26). A blood sample was taken for automated DNA extraction. A small section of each biopsy was formalin fixed and histopathological classification, dysplasia by the Vienna classification (27) and proportion of adenomatous material were determined. For the latter, the percentage of epithelial adenoma nuclei was determined in relation to the total number of nuclei comprising adenoma, non-neoplastic crypts, stroma / lamina propria / muscularis mucosae / submucosa, lymphoid and inflammatory cells. The remainder of each biopsy was snap frozen with liquid nitrogen and stored at -80°C until DNA was extracted using the phenol/chloroform method. A potential limitation in sample characterization was that we could not confirm whether sections used for histopathology were representative of the rest of each biopsy.

Whole Exome Sequencing

Whole exome sequencing of adenoma and matched blood DNA was performed to a mean depth of coverage of 100x at the Beijing Genomics Institute, Hong Kong, using the SureSelect Human 50Mb capture kit (Agilent) and Illumina platforms. A potential limitation of the chosen depth of coverage is failure to detect somatic variants occurring at very low frequency due to tumor heterogeneity.

Bioinformatic Analysis and Identification of Somatic Single Nucleotide Variants (SNVs).

Details of variant calling can be found in the suppl methods.

Validation of Somatic Mutations

Putative protein changing somatic mutations were validated by PCR and Sanger sequencing of original adenoma DNA samples. When the sequencing depth in a matched blood sample was 20x or less, PCR and Sanger sequencing was also performed on the blood DNA sample. Primers were purchased from Eurofins and PCR was completed as described in the supplementary methods.

Identification and Analysis of Recurrently Mutated Genes

Recurrently mutated genes were defined as those with ≥2 validated somatic protein changing mutations in the 20 duodenal adenoma exomes. Data for adenomas 37A1 and 37A4 and for adenomas 24A3 and 24A8 were merged as each of these pairs shared a significant proportion of confirmed somatic mutations indicating that they were not independent tumors. Mutations present in each of these pairs were counted only once. To determine which genes were significantly mutated, all validated variants were analyzed using MutSig v1.0 (http://www.broadinstitute.org/cancer/cga/mutsig). To adjust for

multiple testing and reduce the false discovery rate, q values were calculated (32). Genes with *P*< 0.05 (Fishers Exact test) and a Q<= 0.1 were reported as significantly mutated (see supplementary methods for details).

In order to gain insight into potential mechanisms of tumorigenesis, pathway enrichment analysis was undertaken on all 941 validated somatic mutations using ConsensusPathDB (33) (suppl methods).

Sanger Sequencing in Additional Adenomas

Sanger sequencing of 42 additional adenoma biopsies was used to extend data on somatic mutations in *ERBB3, KRAS, PLCL1, PTCHD2 and WTX* and of 49 additional adenomas for *APC* exon 15 (for details see supplementary methods).

Loss of Heterozygosity (LOH) Analysis

LOH analysis at the *APC* locus was performed on adenomas in which somatic *APC* mutations were not identified by sequencing (details in supplementary methods). A 50% or greater reduction in an allele relative to constitutional DNA was reported as allelic loss.

Identification and Confirmation of Somatic Copy Number Variants (CNVs)

Somatic CNVs were identified by aCGH of 47 duodenal adenomas, 26 from FAP patients and 21 from MAP patients, and matched blood DNA using the BlueGnome CytoChip ISCA 8x60k (v2.0) array (GRCh37) (supplementary methods). Slides were scanned at 3μm resolution and data were analyzed using CytoGenomics software (Agilent). Each putative CNV was confirmed by either independent aCGH analysis using the Illumina CytoSNP-850k v1.0 chip and data analysis with BlueFuse Multi v3.3 or by quantitative (qPCR) using the 7500 Real-Time PCR system (Applied Biosystems) (supplementary methods). CNVs found by aCGH in samples that had been exome sequenced were also validated from exome data using ExomeCNV software (34) (supplementary methods).

Published Data on Somatic APC Mutations in MAP and FAP Adenomas

We compiled a database of somatic *APC* mutations reported in FAP or MAP duodenal or colorectal adenomas via a literature search in PubMed and Google using the search terms 'duodenum', 'colorectum', 'FAP', 'MAP' and 'adenoma'.

Statistical Analysis

Statistical analysis was performed using R (version 3.0.2). The Student's t-test was employed to compare the frequencies of SNVs in FAP and MAP adenomas and Fisher's exact test to compare the frequencies of G>T transversions. A *P* value of less than 0.05 was considered statistically

significant. Correlation of adenoma size with number of SNVs and Spigelman stage with number of SNVs was analyzed by Pearson's correlation coefficient where 1 is a perfect positive correlation, 0 is no correlation and -1 is a perfect negative correlation

Results

Characterization of Patients and Adenomas

Biopsies of 72 apparently independent polyps were obtained (1 to 7 biopsied polyps per patient). Histology confirmed that 69 were adenomas including 42 from 16 patients with FAP and 27 from 10 patients with MAP (Table 1). Two biopsies contained only normal mucosa and one only inflamed ampullary tissue. MAP patients were significantly older than those with FAP (mean 55.0 years versus 42.9 years, $P = 0.006$), but had significantly lower Spigelman stage disease (mode stage II versus IV, $P = 0.031$). Spigelman stage was also lower in MAP than FAP patients from whom adenomas were used for whole exome sequencing (stages II,II,II,II,III vs III,III,III,IV respectively). There was no significant difference in the size of biopsied adenomas from FAP and MAP patients (mean 6.93 mm, range 1-30 mm, SD 6.35 mm versus mean 8.12 mm, range 1.5-25 mm, SD 6.14 mm, $P = 0.4255$) or in the size of FAP and MAP adenomas used for whole exome sequencing (mean 11.1 mm, range 2-25 mm, SD 7.5 mm versus mean 11.7 mm, range 3-25 mm, SD 8.26 mm respectively, $P = 0.867$). All adenomas showed only low grade dysplasia and most had tubular morphology with 7/42 (17%) of FAP adenomas and 2/27

(7%) of MAP adenomas having a villous component (Table 1). The lower Spigelman grade of duodenal disease in MAP than FAP patients reflected smaller adenoma numbers and less frequent villous morphology.

Somatic Mutation Landscape in FAP and MAP Duodenal Adenomas

Whole exome sequencing of 20 duodenal adenomas, 10 from 4 patients with FAP and 10 from 5 patients with MAP, together with matched blood DNA identified 1449 putative protein altering somatic mutations. PCR and Sanger sequencing validated 941 of these (65%, Supplementary Table 1 and 2) including 28 *APC* mutations that were identified initially by manual inspection of the exome data and 913 variants in other genes. Eighty three percent of the validated mutations were nonsynonymous (missense) changes, 13% were stopgains, 2% were splice site mutations, 1% were frameshifts and one was a stoploss. There were significantly more validated protein changing somatic mutations in MAP relative to FAP adenomas (mean 71.6, SD 53.56, range 8-167 *vs* mean 22.5, SD 13.25, range 1-44, P = 0.0115; t-test) (Suppl Figure 1 and Supplementary Table 1). This equated to a mean of 1.43 validated protein changing mutations per Mb in MAP adenoma exomes compared to a mean of 0.44 per Mb in FAP adenoma exomes (Figure 1). The per-Mb rates of protein changing mutations were broadly comparable to those reported previously in non-hypermutated colorectal cancers (21) with MAP duodenal adenomas being towards the top end of the reported range and FAP duodenal adenomas towards the bottom. However, differences in sequencing and variant calling methods demand caution in such comparisons. The

proportion of truncating mutations was also significantly higher in MAP than FAP adenomas ($P = 0.006$). Of 716 mutations in MAP adenomas 481 (67%) were G>T transversions compared to 28/225 (12%) in FAP adenomas ($P <$ 2.2e-16; Fisher's exact test), a finding consistent with failure of base excision repair to remove adenine bases mis-incorporated opposite 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxodG) in MAP adenomas. Pathway enrichment analysis of all validated mutated genes using ConsensusPathDB highlighted over-representation of gene sets involving ECM-receptor interaction networks (q=0.0125), ERBB (q=0.0125), BDNF (q=0.0174), PI3K/AKT (q=0.0287), EGF and FGF (q=0.0414) signaling pathways in FAP adenomas as well as significant enrichment for protein complexes that are part of canonical WNT (q=0.00516) and MAPK (q=0.00516) signaling cascades.

In MAP adenomas, interrogation for protein complex-based sets showed an enrichment for epigenetic transcription regulators (q=0.00263) as well as molecules important in DNA repair pathways (q=0.031) and, consequently, over-representation of genes involved in the maintenance of DNA integrity. The number of mutations in different adenomas from the same individual varied greatly (Suppl. Figure 1).

We also tested for a correlation between adenoma size and the number of confirmed somatic mutations. Although larger adenomas contained more mutations, this did not reach significance for either FAP adenomas (Pearson's product-moment correlation, r = 0.62, $P = 0.054$) or MAP adenomas (r = 0.36, $P= 0.303$).

Despite appearing to be distinct at endoscopy, MAP adenomas 37A1 and 37A4 shared the same somatic *APC* mutations and 30 other validated

somatic variants. A further 167 validated variants were not shared. MAP adenomas 24A3 and 24A8 also appeared distinct at endoscopy but shared the same somatic *APC* mutations and 60 other validated variants while 34 validated variants were not shared. The proportions of adenomatous nuclei also differed between adenomas in these pairs (Table 1). Each pair was considered likely to have diverged from a single progenitor lesion and variants in each pair were counted only once in analyses to identify recurrently mutated genes.

Recurrently Mutated Genes

Sixty-two genes were mutated recurrently in the adenomas subject to whole exome sequencing (Supplementary Table 3) but analysis with MutSig v1.0, which evaluates the number of mutations observed in the context of gene size and the background mutation rate, showed that only 15 were mutated significantly more often than expected (Table 2). Of these, 12 were also mutated significantly in the COSMIC database of somatic mutations in cancer (http://cancer.sanger.ac.uk/cosmic) (Table 2). Truncating mutations were observed recurrently in *APC, PIGA, TRPM1* and *SYNE1* but only *APC* and *PIGA* were mutated significantly above the expected background rate. *PIGA* was not mutated significantly in COSMIC and therefore does not appear to be a driver gene in more extensively studied tumor types.

Extended Analysis of APC, KRAS, PTCHD2, ERBB3, PLCL1 and WTX

To gain further insight into the frequencies and nature of mutations affecting examples of both established and novel candidate driver genes, we extended the analysis of *APC* (in 49 further duodenal adenomas) and *KRAS*, *PTCHD2*, *ERBB3* and *PLCL1* (in 42 further duodenal adenomas) by Sanger sequencing. *PLCL1* was not significantly mutated according to MutSig v1.0, but the 4 *PLCL1* mutations identified during exome sequencing clustered within a region spanning residues 440-547 and this clustering was significant ((35) ($P = 0.004$). Although whole exome sequencing did not identify any mutations in *WTX* it was identified recently as a frequently mutated gene in FAP and MAP colorectal adenomas (20) and is also among the most frequently mutated genes in non-hypermutated colorectal cancer (21). We therefore also sequenced *WTX* in 42 further duodenal adenomas.

Forty further *APC* mutations were identified by Sanger sequencing (Tables 1 and 4) and LOH analysis revealed somatic loss affecting 3 further *APC* alleles in which sequencing was normal. As aCGH detected no CNVs at the *APC* locus the LOH appeared to be copy neutral.

The somatic *APC* mutations and those reported in previous studies of FAP duodenal adenomas (see Supplementary Table 4) clustered 3' to the third (last) β-catenin binding 20 amino acid repeat. This non-random clustering was highly significant ($P = 9.11 \times 10^{-10}$ by the method of Ye et al (35)) and different to the clustering of somatic *APC* mutations in FAP-associated colorectal adenomas (Supplementary Table 4) that occurs after the first and second 20 amino acid repeats ($P < 3.72 \times 10^{-16}$ and $P < 3.88 \times 10^{-29}$). In FAP duodenal adenomas, 15 of the 30 *APC* mutations we identified were insertion of an A in

the A_6 tract at codons 1554-6 (c.4659dupA; E1554fsX5). This mutation also accounted for 17/35 previously reported somatic *APC* mutations in FAP duodenal tumors but only 1/296 in FAP colorectal adenomas (Supplementary Table 4, $P < 0.0001$; Fisher's exact test).

In MAP duodenal adenomas where biallelic *APC* mutations were identified, significant clustering occurred between codons 1530 and 1576 ($P = 1.25 \times 10^{-7}$) despite the presence of GAA sequences throughout the coding region that could be mutated to stop codons by G>T transversion with only one instance of E1554fsX5 observed (in the adenoma pair 37A1 and 37A4, Supplementary Table 4).

We did not observe any somatic *WTX* mutations in 60 independent duodenal adenomas (Table 3). This was significantly different ($P = 0.0038$, Fisher's exact test) to the findings reported by Rashid et al. (20) in FAP and MAP colorectal adenomas, where 17 truncating mutations were identified in 128 adenomas, making *WTX* the most frequently mutated gene after *APC*. WTX forms a complex with APC, Axin and β-TrCP2 that degrades β-catenin. It is likely that the differences we observed between duodenal and colorectal adenomas in the positions or presence of *APC* and *WTX* mutations reflect different requirements for β-catenin signaling for tumorigenesis in these contexts.

After *APC, KRAS* was the most frequently mutated gene in duodenal adenomas (12/60, 20%) and *KRAS* mutations were significantly more frequent in MAP than FAP adenomas (8/22 vs 4/38, $P < 0.023$, Fisher's exact

test). Only 3/8 *KRAS* mutations in MAP duodenal adenomas were the c.34 G:C>T:A (G12C) mutation that has been considered as a potential biomarker of MAP in patients with multiple colorectal adenomas (36). MAP patients whose adenomas harbored *KRAS* mutations appeared to have lower Spigelman stage polyposis than corresponding FAP patients (stages II,II,II,II,III in MAP vs II,IV,IV in FAP, Table 1).

Six somatic *PTCHD2* mutations were identified in 60 independent adenomas, 3 by whole exome sequencing and 3 by sequencing of additional adenomas. Five had CADD scores above 20 (i.e. corresponding to the top 1% of substitutions in terms of predicted deleterious effects). Adenomas 3A2 and 37A1 each contained two *PTCHD2* mutations but one of those in 37A1 was unlikely to be of functional significance (Supplementary Table 5). Six independent *PLCL1* mutations were also observed: 4 in whole exomes and 2 following targeted sequencing. The latter 2 did not cluster with the others (Supplementary Table 5). All but one of the *PLCL1* mutations had CADD scores above 20. No further mutations of *ERBB3* were identified by analysis of the 42 additional adenomas but the 2 mutations identified during exome sequencing had CADD scores of 28.4 and 30 and are very likely to impact function (Supplementary Table 5).

Array CGH

Array CGH revealed 8 CNVs (5 losses and 3 gains) in 5 of 19 MAP duodenal adenomas (Table 4) compared to none in 26 FAP adenomas ($P = 0.0052$,

Fisher's exact test). All were confirmed by either quantitative PCR or by using a second array, the Illumina CytoSNP-850k v1.0. Several involved genes in the BMP/TGF-β signaling pathway: the deletion at 18q21.1 in adenoma 44A4 included *SMAD4* and that at 9q22 included *ENG*, while the 15q11.1-15q21.1 gains in adenomas 23A3 and 23A4 included *GREM1*, a BMP antagonist.

Discussion

Duodenal polyposis and cancer present a major challenge in the clinical management of FAP and MAP, but remain understudied and poorly understood. This study is the first to characterize comprehensively the burden and pattern of somatic mutations in duodenal adenomas from patients with FAP or MAP. We found that MAP duodenal adenomas carried a significantly higher burden of somatic protein changing mutations, truncating mutations and CNVs than FAP duodenal adenomas even though MAP patients had lower Spigelman stage duodenal polyposis than FAP patients. The greater mutation burden in MAP adenomas appears to reflect defective base excision repair. Although longitudinal or prospective studies of duodenal polyposis in MAP have not been reported, case reports have highlighted the occurrence of duodenal cancer in MAP patients in the absence of advanced duodenal polyposis (3,8). These observations and our data suggest that current recommendations to manage MAP duodenal polyposis using Spigelman staging in the same way as for FAP (1,6) may not be appropriate. A low polyp count in a patient with MAP may be falsely reassuring and, in addition, we did not find a significant correlation between adenoma size and mutation burden.

Mutation burdens in some small MAP adenomas were among the highest we observed. Large, prospective clinical studies could provide a better evidence base for duodenal surveillance recommendations and intervention in MAP.

Our data confirm the importance of *APC* and *KRAS* mutations as drivers of duodenal tumorigenesis in FAP and MAP but show that in contrast with the colorectum (20,21,37,38) *WTX* is not a significant driver gene in early duodenal tumorigenesis. Neither did we identify by exome sequencing any mutations in a number of known driver genes including *NRAS, CTNNB1, FBXW7 and TP53* that were mutated recurrently in previous studies of sporadic or FAP-associated colorectal adenomas (37,39) and that are also mutated in sporadic duodenal adenocarcinomas (40,41). They may be mutated later in duodenal tumorigenesis.

The somatic *APC* mutations we identified in FAP and MAP duodenal adenomas clustered 3' to the mutation cluster region observed in FAP-associated and sporadic colorectal adenomas and cancers. Groves et al (19) and Miyaki et al (42) have reported similar findings. These more 3' mutations are predicted to lead to truncated APC proteins that retain 3 β-catenin binding 20 AA repeats in the majority of duodenal tumors rather than either 1 or 2 repeats as occurs in colorectal tumors. In FAP duodenal adenomas we found that 14/25 (56%) somatic *APC* mutations were ins A mutations at codons 1554-6 (4661 G>GA c.4659dupA; E1554fxs4). This is consistent with data we compiled from previous reports in which this mutation accounted for 17/35 mutations (49%). Although very uncommon in FAP colorectal adenomas

(1/296 mutations in the reports we identified, Supplementary Table 4), this mutation has been seen recurrently in colorectal adenomas from patients with attenuated FAP (43–45) where it appears to occur as a "third hit" further reducing the activity of the attenuated germline mutant allele. We did not find any evidence for third hits affecting *APC* in duodenal adenomas. Instead, this change and the others clustering after the 3rd 20 AA repeat are likely to be selected for as second hits in duodenal tumorigenesis because they determine a specific level of β-catenin signaling that is lower than that selected for in colorectal tumorigenesis. A subtly different β-catenin signaling requirement in duodenal adenomas may also explain the absence of *WTX* mutations.

In addition to *APC* and *KRAS*, 10 of the 13 other genes that were mutated significantly upon whole exome sequencing of duodenal adenomas are also mutated significantly in the COSMIC database of somatic mutations in cancer (Table 2). These genes are likely to be drivers in FAP and MAP duodenal tumors as well as in other tumor types. Following whole exome sequencing we investigated the recurrently mutated genes *PTCHD2, ERBB3* and *PLCL1* in a set of 42 additional duodenal adenomas. We identified further mutations in *PTCHD*2 (N = 3) and *PLCL1* (N = 2), supporting a role for these genes as drivers in duodenal tumorigenesis. *PLCL1* encodes a multivalent adaptor protein (46). Four of six mutations identified in this study were missense changes clustered around the X-Box region of the PLC core domain. A truncating mutation of *PLCL1* (S931X) was also identified in 1 of 14 colorectal adenoma exomes in the study of Rashid et al (20). *PTCHD2* (*DISP3*) has

been assigned to the family of Patched-domain containing receptors based on *in silico* characterization and is likely involved in Hedgehog signaling (47).

A number of genes such as *MLL3* and *ATRNL1* in which we identified only single truncating mutations were also mutated recurrently in FAP and/or MAP colorectal adenomas in other recent studies (20). They represent candidate driver genes in duodenal as well as colorectal tumorigenesis. aCGH identified CNVs exclusively in MAP duodenal adenomas and several included genes (*SMAD4*, *ENG* and *GREM1*) that regulate BMP signaling and have established roles in gastrointestinal cancer. aCGH lacks sensitivity in the context of heterogeneous tumor samples that comprise a mixture of neoplastic and non-neoplastic cells and we are likely to have underestimated the true frequency of CNVs. Pathway enrichment analysis of all validated mutations provided an approach to evaluate the potential roles of multiple genes with related functions. It highlighted involvement of Wnt, ERBB, PI3K/AKT, EGF, FGF and ECM-receptor signaling in FAP adenomas and of DNA repair pathways and epigenetic transcription regulators in MAP adenomas. Dysregulation of these pathways is well established in tumorigenesis and they are targets for drugs in clinical use or under development. So far only EGF signaling has been targeted in clinical trials for duodenal polyposis (16). Our data point to additional and novel opportunities for intervention but they also highlight the molecular genetic heterogeneity of duodenal adenomas. Only genes that regulate the Wnt pathway were mutated consistently. The highly specific and restricted pattern of *APC* mutation and the absence of *WTX* mutation that we observed in duodenal adenomas

suggest that a narrow range of β-catenin activity may be required for duodenal tumorigenesis. Therapeutic manipulation of this activity may hold particular promise for prevention and treatment.

Acknowledgements:

Special thanks to Dr Fiona Lalloo and Mr Jim Hill for assistance with communication with participating patients.

References

1. Vasen HFA, Möslein G, Alonso A, Aretz S, Bernstein I, Bertario L, et al. Guidelines for the clinical management of familial adenomatous polyposis (FAP). Gut. 2008;57:704–13.
2. Bülow S, Christensen IJ, Højen H, Björk J, Elmberg M, Järvinen H, et al. Duodenal surveillance improves the prognosis after duodenal cancer in familial adenomatous polyposis. Color Dis. 2012;14:947–52.
3. Vogt S, Jones N, Christian D, Engel C, Nielsen M, Kaufmann A, et al. Expanded Extracolonic Tumor Spectrum in MUTYH-Associated Polyposis. Gastroenterology. 2009;137:1976–1985.e10.
4. Walton S-J, Kallenberg FGJ, Clark SK, Dekker E, Latchford A. Frequency and Features of Duodenal Adenomas in Patients With MUTYH-Associated Polyposis. Clin Gastroenterol Hepatol. 2016;14:986–92.
5. Spigelman AD, Williams CB, Talbot IC, Domizio P, Phillips RK. Upper gastrointestinal cancer in patients with familial adenomatous polyposis. Lancet (London, England). 1989;2:783–5.
6. Syngal S, Brand RE, Church JM, Giardiello FM, Hampel HL, Burt RW, et al. ACG Clinical Guideline: Genetic Testing and Management of Hereditary Gastrointestinal Cancer Syndromes. Am J Gastroenterol. 2015;110:223–62.
7. Serrano PE, Grant RC, Berk TC, Kim D, Al-Ali H, Cohen Z, et al. Progression and Management of Duodenal Neoplasia in Familial Adenomatous Polyposis. Ann Surg. 2015;261:1138–44.
8. Nielsen M, Poley JW, Verhoef S, van Puijenbroek M, Weiss MM, Burger

GT, et al. Duodenal carcinoma in MUTYH-associated polyposis. J Clin Pathol. BMJ Group; 2006;59:1212–5.

9. Alderlieste YA, Bastiaansen BA, Mathus-Vliegen EMH, Gouma DJ, Dekker E. High rate of recurrent adenomatosis during endoscopic surveillance after duodenectomy in patients with familial adenomatous polyposis. Fam Cancer. Springer Netherlands; 2013;12:699–706.
10. Ma T, Jang EJ, Zukerberg LR, Odze R, Gala MK, Kelsey PB, et al. Recurrences are common after endoscopic ampullectomy for adenoma in the familial adenomatous polyposis (FAP) syndrome. Surg Endosc. 2014;28:2349–56.
11. Heiskanen I, Kellokumpu I, Järvinen H. Management of duodenal adenomas in 98 patients with familial adenomatous polyposis. Endoscopy. 1999;31:412–6.
12. Steinbach G, Lynch PM, Phillips RKS, Wallace MH, Hawk E, Gordon GB, et al. The Effect of Celecoxib, a Cyclooxygenase-2 Inhibitor, in Familial Adenomatous Polyposis. N Engl J Med. 2000;342:1946–52.
13. Giardiello FM, Yang VW, Hylind LM, Krush AJ, Petersen GM, Trimbath JD, et al. Primary Chemoprevention of Familial Adenomatous Polyposis with Sulindac. N Engl J Med. 2002;346:1054–9.
14. Phillips RKS, Wallace MH, Lynch PM, Hawk E, Gordon GB, Saunders BP, et al. A randomised, double blind, placebo controlled study of celecoxib, a selective cyclooxygenase 2 inhibitor, on duodenal polyposis in familial adenomatous polyposis. Gut. 2002;50:857–60.
15. Arber N, Eagle CJ, Spicak J, Rácz I, Dite P, Hajer J, et al. Celecoxib for the Prevention of Colorectal Adenomatous Polyps. N Engl J Med.

Massachusetts Medical Society ; 2006;355:885–95.

16. Samadder NJ, Neklason DW, Boucher KM, Byrne KR, Kanth P, Samowitz W, et al. Effect of Sulindac and Erlotinib vs Placebo on Duodenal Neoplasia in Familial Adenomatous Polyposis. JAMA. 2016;315:1266.
17. Lamlum H, Ilyas M, Rowan A, Clark S, Johnson V, Bell J, et al. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. Nat Med. 1999;5:1071–5.
18. Albuquerque C, Breukel C, van der Luijt R, Fidalgo P, Lage P, Slors FJM, et al. The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. Hum Mol Genet. 2002;11:1549–60.
19. Groves C, Lamlum H, Crabtree M, Williamson J, Taylor C, Bass S, et al. Mutation cluster region, association between germline and somatic mutations and genotype-phenotype correlation in upper gastrointestinal familial adenomatous polyposis. Am J Pathol. American Society for Investigative Pathology; 2002;160:2055–61.
20. Rashid M, Fischer A, Wilson CH, Tiffen J, Rust AG, Stevens P, et al. Adenoma development in familial adenomatous polyposis and MUTYH-associated polyposis: somatic landscape and driver genes. J Pathol. Wiley-Blackwell; 2016;238:98–108.
21. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487:330–7.

22. Gallinger S, Vivona AA, Odze RD, Mitri A, O'Beirne CP, Berk TC, et al. Somatic APC and K-ras codon 12 mutations in periampullary adenomas and carcinomas from familial adenomatous polyposis patients. Oncogene. 1995;10:1875–8.

23. Kashiwagi H, Spigelman AD, Talbot IC, Debinski HS, McKie AB, Lemoine NR, et al. p53 and K-ras status in duodenal adenomas in familial adenomatous polyposis. Br J Surg. 1997;84:826–9.

24. Norheim Andersen S, Løvig T, Fausa O, Rognum TO. Germline and somatic mutations in exon 15 of the APC gene and K-ras mutations in duodenal adenomas in patients with familial adenomatous polyposis. Scand J Gastroenterol. 1999;34:611–7.

25. Wagner PL, Chen Y-T, Yantiss RK. Immunohistochemical and Molecular Features of Sporadic and FAP-associated Duodenal Adenomas of the Ampullary and Nonampullary Mucosa. Am J Surg Pathol. 2008;32:1388–95.

26. Saurin J-C, Gutknecht C, Napoleon B, Chavaillon A, Ecochard R, Scoazec J-Y, et al. Surveillance of Duodenal Adenomas in Familial Adenomatous Polyposis Reveals High Cumulative Risk of Advanced Disease. J Clin Oncol. 2003;22:493–8.

27. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut. BMJ Group; 2000;47:251–5.

28. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.

29. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

30. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.

31. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics. 2012;28:311–7.

32. Storey JD. The positive false discovery rate: a bayesian interpretation and the q-value. Ann Stat. 2003;31:2013–35.

33. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. Nat Protoc. Nature Research; 2016;11:1889–907.

34. Sathirapongsasuti JF, Lee H, Horst BAJ, Brunner G, Cochran AJ, Binder S, et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics. 2011;27:2648–54.

35. Ye J, Pavlicek A, Lunney EA, Rejto PA, Teng C-H. Statistical method on nonrandom clustering with application to somatic mutations in cancer. BMC Bioinformatics. 2010;11:11.

36. Nielsen M, Morreau H, Vasen HFA, Hes FJ. MUTYH-associated polyposis (MAP). Crit Rev Oncol Hematol. 2011;79:1–16.

37. Nikolaev SI, Sotiriou SK, Pateras IS, Santoni F, Sougioultzis S, Edgren

H, et al. A Single-Nucleotide Substitution Mutator Phenotype Revealed by Exome Sequencing of Human Colon Adenomas. Cancer Res. 2012;72:6279–89.

38. Sanz-Pamplona R, Lopez-Doriga A, Pare-Brunet L, Lazaro K, Bellido F, Alonso MH, et al. Exome Sequencing Reveals AMER1 as a Frequently Mutated Gene in Colorectal Cancer. Clin Cancer Res. 2015;21:4709–18.

39. Borras E, San Lucas FA, Chang K, Zhou R, Masand G, Fowler J, et al. Genomic Landscape of Colorectal Mucosa and Adenomas. Cancer Prev Res (Phila). 2016;9:417–27.

40. Laforest A, Aparicio T, Zaanan A, Silva FP, Didelot A, Desbeaux A, et al. ERBB2 gene as a potential therapeutic target in small bowel adenocarcinoma. Eur J Cancer. 2014;50:1740–6.

41. Yuan W, Zhang Z, Dai B, Wei Q, Liu J, Liu Y, et al. Whole-exome sequencing of duodenal adenocarcinoma identifies recurrent Wnt/beta-catenin signaling pathway mutations. Cancer. 2016;122:1689–96.

42. Miyaki M, Yamaguchi T, Iijima T, Takahashi K, Matsumoto H, Yasutome M, et al. Difference in characteristics of APC mutations between colonic and extracolonic tumors of FAP patients: Variations with phenotype. Int J Cancer. 2008;122:2491–7.

43. Spirio LN, Samowitz W, Robertson J, Robertson M, Burt RW, Leppert M, et al. Alleles of APC modulate the frequency and classes of mutations that lead to colon polyps. Nat Genet. 1998;20:385–8.

44. Su L-K, Barnes CJ, Yao W, Qi Y, Lynch PM, Steinbach G. Inactivation of Germline Mutant APC Alleles by Attenuated Somatic Mutations: A

Molecular Genetic Mechanism for Attenuated Familial Adenomatous Polyposis. Am J Hum Genet. 2000;67:582–90.

45. Sieber OM, Segditsas S, Knudsen AL, Zhang J, Luz J, Rowan AJ, et al. Disease severity and genetic pathways in attenuated familial adenomatous polyposis vary greatly but depend on the site of the germline mutation. Gut. BMJ Publishing Group Ltd and British Society of Gastroenterology; 2006;55:1440–8.

46. Sugiyama G, Takeuchi H, Kanematsu T, Gao J, Matsuda M, Hirata M. Phospholipase C-related but catalytically inactive protein, PRIP as a scaffolding protein for phospho-regulation. Adv Biol Regul. 2013;53:331–40.

47. Katoh Y, Katoh M. Identification and characterization of DISP3 gene in silico. Int J Oncol. 2005;26:551–6.

Figure Legend

Figure 1. Box plot showing per megabase (Mb), median and 25th and 75th percentiles and range of confirmed non-synonymous SNVs in FAP and MAP duodenal adenomas.

Tables and Figures

Table 1. Details of patients and adenomas studied. Adenomas with IDs in gray boxes were subject to whole exome sequencing. TA = tubular adenoma, TVA = tubulovillous adenoma, VA = villous adenoma, LGD = low grade dysplasia. Non-independent adenomas are shown in bold. Total numbers of adenomas were counted following chromoendoscopy.

Patient	Adenoma	Germline Mutation	Somatic *APC* mutation(s)		FAP/ MAP	Age (years)	Patient Sex	Total Number of Adenomas*	Size of adenoma (mm)	Histology	% Adeno-matous Tissue	Spigel-man Stage
2	2A5	c.994C>T; p.R332X	None identified		FAP	43	M	1	2	TA LGD	40%	I
3	3A2	*APC* Promoter-Exon 2 Deletion	None identified		FAP	38	F	38	5	TA LGD	40%	III
	3A3		4645 C>T	Q1549X					5	TA LGD	90%	
	3A4		4659dupA	E1554fsX5					6	TA LGD	30%	
4	4A1	*APC* Promoter-Exon 2 Deletion	None identified		FAP	69	M	21	3	TA LGD	50%	II
	4A2		None identified						3	TA LGD	50%	
	4A3		None identified						4	TA LGD	50%	
8	8A2	c.477C>G; p.Y159X	4348 C>T	R1450X	FAP	36	M	6	2	TA LGD	40%	II
	8A3		4645 C>T	Q1549X					3	TA LGD	50%	
17	17A1	*APC* Exon 4_5 Deletion	4612_4613delGA	E1538IfsX5	FAP	38	F	79	15	TA LGD	80%	IV
	17A2		4691 T>G	L1564X					20	TVA LGD	80%	
	17A3		None identified						5	TA LGD	40%	
19	19A1	c.2805C>A; p.Y935X	4659dupA	E1554fsX5	FAP	35	M	61	15	TVA LGD	10%	IV
	19A2		4659dupA	E1554fsX5					8	TA LGD	30%	
	19A3		4659dupA	E1554fsX5					10	TA LGD	20%	
	19A4		4729 G>T	E1577X					12	TA LGD	5%	
21	21A1	c.3785 dupA; p.Y1262FfsX2	4659dupA	E1554fsX5	FAP	41	F	110	30	TVA LGD	80%	IV
	21A3		4659dupA	E1554fsX5					5	TA LGD	20%	
22	22A2	c.3366_69delTCAA; p.N1122fsX2	4659dupA	E1554fsX5	FAP	32	F	8	2	TA LGD	30%	II
	22A3		None identified						4	TA LGD	10%	
	22A4		4381G>T	E1461X					4	TA LGD	60%	
23	23A1	c.477C>G; p.Y159X	4659dupA	E1554fsX5	FAP	45	F	16	2	TA LGD	10%	II
	23A2		4659dupA	E1554fsX5					4	TA LGD	60%	
29	29A2	c.3203_3205delTCAA; p.S1068fsX56	4698 del23bp	D1566fsX16	FAP	53	M	14	4	TA LGD	5%	IV
	29A3		4659dupA	E1554fsX5					4	VA LGD	40%	
	29A4		4659dupA	E1554fsX5					5	TVA LGD	35%	
30	30A1	c.3198 ACAAT>CAAT; p.R1067fsX59	4659dupA	E1554fsX5	FAP	49	F	26	4	TA LGD	50%	III
	30A3		4592dupA	N1531KfsX2					8	TA LGD	40%	
50	50A1	c.637C>T; p.Arg213X	None identified		FAP	31	F	2	2	TA LGD	50%	I
	50A2		None identified						1	TA LGD	50%	
51	51A1	c.637 C>T; p.R213X	4606 G>T	E1536X	FAP	42	M	17	10	TA LGD	50%	III
	51A3		4659dupA	E1554fsX5					4	TA LGD	30%	
	51A4		LOH (nt5037)						2	TA LGD	20%	

	51A5		None identified						2	TA LGD	5%	
52	52A1	c.3863 GA>A; p.G1288fsX16	3862delG	G1288fsX17	FAP	37	M	5	6	TA LGD	50%	IV
	52A2		4734 T>A	C1578X					8	TA LGD	50%	
	52A3		4659dupA	E1554fsX5					25	TVA LGD	80%	
	52A4		4348 C>T	R1450X					15	TA LGD	60%	
	52A5		4393_4394dupAG	S1465RfsX9					6	TVA LGD	60%	
D1	D1A1	c.3203-3205delTCAA	None identified		FAP	56	M	20	3	TA LGD	30%	II
	D1A2		None identified						3	TA LGD	20%	
D4	D4A3	c.3176_3180delAAATA; p.I1060TfsX3	4722_4725delACTA	I1574MfsX75	FAP	29	M	10	10	TA LGD	60%	II
7	7A2	c.536A>G; p.Y179C (HOM)	4735 G>T	E1560X	MAP	62	F	1	1,5	TALGD	40%	I
			None identified									
24	24A1	c.536 A>G; p.Y179C (HOM)	4678 G>T	E1560X	MAP	59	M	15	15	TA LGD	40%	III
			LOH (rs2019720)									
	24A2		4381 G>T	E1461X					10	TA LGD	40%	
			4654 G>T	E1552X								
	24A3		3502 G>T	E1168X					15	TA LGD	50%	
			4654 G>T	E1552X								
	24A4		3502 G>T	E1168X					8	TA LGD	50%	
			4654 G>T	E1552X								
	24A5		2281 G>T	E761X					6	TA LGD	10%	
			None identified									
	24A6		4639 G>T	E1547X					5	TA LGD	20%	
			None identified									
	24A7		2863 G>T	E955X					15	TVA LGD	70%	
			4612G>T	E1538X								
	24A8		3502 G>T	E1168X					12	TVA LGD	90%	
			4654 G>T	E1552X								
26	26A1	c.1438 G>T; p.E480X (HOM)	None identified		MAP	51	F	1	5	TA LGD	60%	II
			4639 G>T	E1547X								
33	33A1	c.1438 G>T; p.E480X (HOM)	None identified		MAP	68	F	2	3	TA LGD	50%	I
	33A2		None identified						3	TA LGD	50%	
36	36A1	c.1438 G>T; p.E480X (HOM)	2962 G>T	E988X	MAP	65	F	3	3	TA LGD	50%	II
			4639 G>T	E1547X								
	36A3		3845 C>A	S1282X					8	TA LGD	50%	
			4726 G>T	E1576X								
37	37A1	c.1214 C>T; p.P405L and c.1187 G>A; p.G396D	526 G>T	E176X	MAP	66	F	2	25	TA LGD	40%	II
			4659dupA	E1554fsX5								
	37A4		526 G>T	E176X					25	TA LGD	70%	
			4659dupA	E1554fsX5								

38	38A1	c.739 T>C; p.R247X and c.536 G>A; p.Y179C	4381G>T	E1461X	MAP	47	M	4	5	TA LGD	30%	II
			None identified									
	38A2		3460 G>T	E1154X					5	TA LGD	30%	
			4381 G>T	E1461X								
	38A3		4381G>T	E1461X					5	TA LGD	50%	
			3460 G>T	E1154X								
	38A5		4381G>T	E1461X					5	TA LGD	50%	
			3460 G>T	E1154X								
39	39A1	c.1438 G>T; p.E480X (HOM)	4639 G>T	E1547X	MAP	49	F	3	9	TA LGD	40%	II
			None identified									
	39A3		LOH (D5S346)						9	TA LGD	40%	
			None identified									
41	41A2	c.1438 G>T; p.E480X (HOM)	None identified		MAP	54	F	2	5	TA LGD	80%	II
			None identified									
	41A3		4729 G>T	E1577X					4	TA LGD	10%	
44	44A1	c.1240 C>T; p.Q414X (HOM)	None identified		MAP	42	M	4	5	TA LGD	60%	II
			4639 G>T	E1547X								
	44A2		2311 G>T	E771X					5	TA LGD	60%	
			4630 G>T	E1544X								
	44A4		4588 G>T	E1530X					4	TA LGD	60%	
			3406 G>T	E1136X								

Table 2. Significantly mutated genes identified by MutSig analysis of mutations in Supplementary Table 3 and COSMIC. Every mutation was assigned a CADD score to assess potential functional impact and deleteriousness (Suppl methods). Variants shaded in light gray were present in only MAP adenomas, variants in white were present only in FAP adenomas and those shaded in dark gray were detected in both FAP and MAP adenomas. An asterisk denotes variants that were identified more than once with a superscript number to designate the number of times the variant was detected.

Rank (#)	Gene	Chr	Genomic location	Ref	Alt	Predicted protein	Number of variants	P value	FDR (q)	P value (COSMIC)	FDR (q) (COSMIC)	Significantly recurrently mutated in COSMIC (P < 0.05 & q <= 0.1)	CADD PHRED
1	*APC*	5	112111429	G	T	E176X	24	4.33E-17	2.68E-15	1.69E-103	3.49E-102	TRUE	41
		5	112173602	G	T	E771X							39
		5	112174253	G	T	E988X							39
		5	112174697	G	T	E1136X							39
		5	112174793	G	T	E1168X							39
		5	112175136	C	A	S1282X							37
		5	112175672	G	T	E1461X							37
		5	112175897	G	T	E1536X							42
		5	112175921	G	T	E1544X							39
		5	112175945	G	T	E1552X							39
		5	112175951 *4	G	GA	E1554fsX5							33
		5	112175969	G	T	E1560X							41
		5	112175982	T	G	L1564X							38
		5	112176017	G	T	E1576X							43
		5	112176025	T	A	C1578X							36
		5	112174751	G	T	E1154X							38
		5	112175879 *2	G	T	E1530X							42
		5	112175930	G	T	E1547X							42
		5	112175879	G	GA	N1531KfsX2							35
		5	112175639	C	T	R1450X							38
2	*PIGA*	X	15343189	C	T	E78K	4	8.16E-05	2.53E-03	7.80E-01	1.00E+00	FALSE	33
		X	15342923	G	T	P116H							43
		X	15342994	C	T	SPLICE							25.9
		X	15349456	AT	A	N199fsX4							30
3	*SLC4A3*	2	220500412 *3	G	A	G691R	3	1.66E-04	3.43E-03	2.38E-07	1.23E-06	TRUE	21.8
4	*KRAS*	12	25398284	C	T	G12D	4	4.79E-04	5.94E-03	0.00E+00	0.00E+00	TRUE	25.3
		12	25398285 *2	C	A	G12C							33
		12	25398285	C	T	G12S							31
5	*OR51T1*	11	4903600	C	A	F157L	2	6.71E-04	5.94E-03	7.75E-06	3.45E-05	TRUE	0.074
		11	4904017	G	T	L296F							25.3
6	*FLG2*	1	152325661	G	T	S1534Y	2	6.71E-04	5.94E-03	7.82E-25	9.70E-24	TRUE	23.2
		1	152329718	G	T	H182N							0.92

7	*RBMXL3*	X	114425545	G	A	R514Q	2	6.71E-04	5.94E-03	8.40E-09	4.73E-08	TRUE	21.2
		X	114424797	G	T	G256C							1.495
8	*TRAM1L1*	4	118005732	C	A	G273V	2	1.98E-03	1.36E-02	4.24E-03	1.31E-02	TRUE	23.5
		4	118005846	A	G	M235T							0.076
9	*KRT5*	12	52914023	C	T	A20T	2	1.98E-03	1.36E-02	7.80E-06	3.45E-05	TRUE	1.001
		12	52910917	C	A	A398S							23.1
10	*SFTPD*	10	81706265	C	T	R50C	2	3.89E-03	2.29E-02	1.64E-02	4.61E-02	TRUE	24.5
		10	81706268	G	A	D51Y							24.8
11	*IGFN1*	1	201181973 *3	C	T	S69F	3	4.07E-03	2.29E-02	1.14E-01	3.06E-01	FALSE	23.3
12	*CYLC1*	X	83128944	G	T	E410X	2	1.29E-02	6.68E-02	7.35E-05	2.68E-04	TRUE	39
		X	83128633	C	A	A306D							0.038
13	*PTCHD2*	1	11561594	G	T	G182S	3	1.85E-02	8.80E-02	1.11E-02	3.28E-02	TRUE	0.018
		1	11584030	G	T	Q798H							22.5
		1	11591019	G	T	C1035F							27.6
14	*ERBB3*	12	56480320	C	A	L143M	2	2.14E-02	8.86E-02	5.36E-05	2.08E-04	TRUE	28.4
		12	56487261	C	G	N469K							30
15	*NONO*	X	70514194	C	A	P156T	2	2.14E-02	8.86E-02	4.45E-01	1.00E+00	FALSE	13.24
		X	70514212	G	A	E162K							33

Table 3. Summary of somatic analyses completed including exome analysis, arrayCGH, *APC* LOH analysis and targeted sequencing of *APC, KRAS, WTX, PTCHD2, ERBB3* and *PLCL1*. * Analysis completed but no mutation identified. ■ Mutation identified by exome sequencing. ◆ LOH or CNV detected. O Mutation detected by targeted sequencing. A blank well denotes where a sample was not analysed. Non-independent adenomas are shown in bold, totals reflect the duplication. Gray shading denotes samples that underwent exome sequencing. The table is split to represent the analyses completed on the FAP adenomas in the first section followed by the MAP adenomas in the lower section. The total number of samples screened and mutations detected is given for both the FAP and MAP adenomas individually and then summed across the whole cohort at the end of the table.

Adenoma Sample	FAP/MAP	Exome Sequencing	*APC* Sequencing	*WTX* Sequencing	*KRAS* Sequencing	*PLCL1* Sequencing	*PTCHD2* Sequencing	*ERBB3* Sequencing	*APC* LOH	ArrayCGH
17A1	FAP		■	*	*	*	*	*	*	*
17A2			■	*	*	■	*	*		*
30A1			■	*	*	*	*	*		*
30A3			■	*	*	*	*	*	*	*
51A1			■	*	*	■	*	*		*
51A3			■	*	*	*	*	*		*
51A4			*	*	*	*	*	*	◆	*
52A2			■	*	*	*	*	*	*	*
52A3			■	*	*	*	*	*		*
52A4			■	*	■	*	*	■		*
2A5			*	*	*	*	*	*	*	
3A2			*	*	*	*	OO	*	*	*
3A3			O	*	*	*	*	*		*
3A4			O	*	*	*	*	*		
4A1			*	*	O	*	*	*	*	*
4A2			*	*	*	*	*	*	*	
4A3			*	*	*	*	*	*	*	*
8A3			O	*	*	O	*	*		*
19A1			O	*	*	*	*	*		*
19A2			O	*	*	*	*	*		*
19A3			O	*	*	*	*	*		*
19A4			O	*	*	*	O	*		*
21A3			O	*	*	*	*	*		
22A2			O	*	*	*	*	*		

22A3			*	*	*	*	*	*	*	
22A4			O	*	*	*	*	*		*
23A1			O	*	*	*	*	*		
23A2			O	*	*	*	*	*		*
29A2			O	*	O	*	*	*		
29A3			O	*	O	*	*	*		*
29A4			O	*	*	*	*	*		
50A1			*	*	*	*	*	*	*	
50A3			*	*	*	*	*	*	*	
51A5			*	*	*	*	*	*	*	
52A1			O	*	*	*	*	*		
D1A1			*	*	*	*	*	*	*	
D1A2			*	*	*	*	*	*	*	
D4A3			*	*	*	*	*	*	*	
8A2			O							*
17A3			*						*	*
21A1			O							*
52A5			O							*
FAP adenomas analysed		10	42	38	38	38	38	38	17	26
FAP adenomas with mutations		n/a	28 (66.6%)	0	4 (10.5)	3 (7.9%)	2 (5.2%)	1 (2.6%)	1 (5.9%)	0

Adenoma Sample	**FAP/MAP**	**Exome Sequencing**	***APC* Sequencing**	***WTX* Sequencing**	***KRAS* Sequencing**	***PLCL1* Sequencing**	***PTCHD2* Sequencing**	***ERBB3* Sequencing**	***APC* LOH**	**ArrayCGH**
24A1	MAP		■	*	*	*	*	*	◆	*
24A3			■■	*	■	*	*	■		◆
24A8			■■	*	■	*	*	■		*
36A1			■■	*	*	*	*	*	*	*
36A3			■■	*	*	*	*	*		*
37A1			■■	*	■	■■	■■	*		◆◆
37A4			■■	*	■	■	*	*		*
38A2			■■	*	■	*	■	*		◆
44A2			■■	*	*	*	*	*		◆◆◆
44A4			■■	*	*	*	*	*		*
7A2			O	*	*	*	*	*		*
24A2			OO	*	*	*	*	*		
24A4			OO	*	O	*	*	*		◆

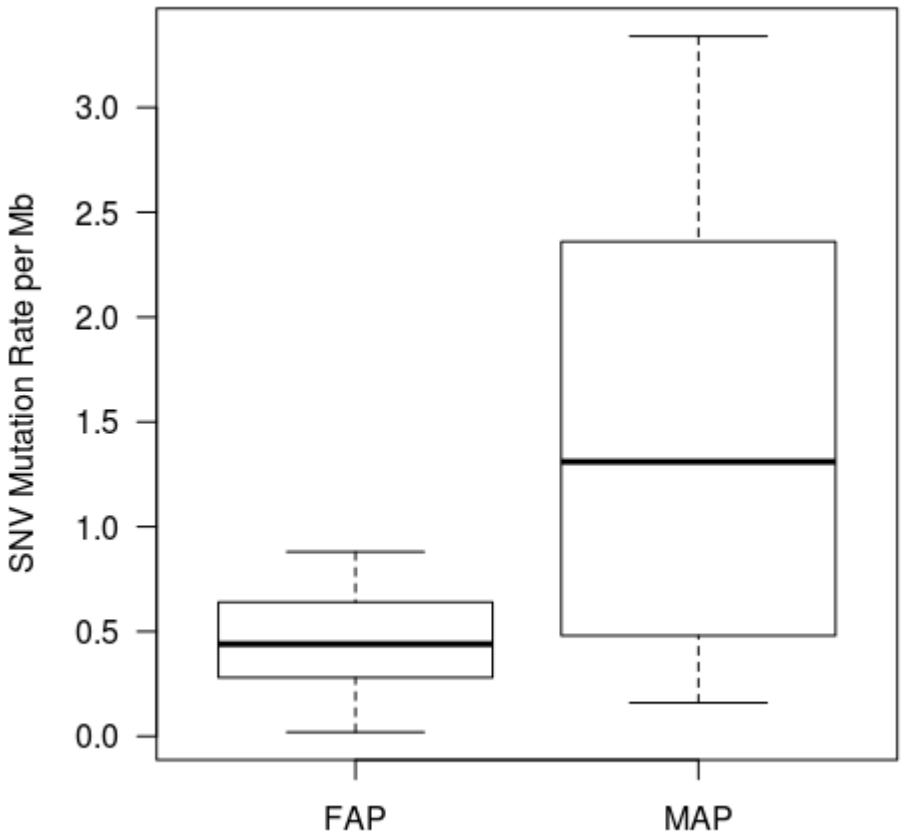
24A5			o	*	*	*	*	*		*
24A7			oo	*	*	*	*	*		
26A1			o	*	*	*	*	*		*
33A1			*	*	*	*	*	*	*	
33A2			*	*	*	*	*	*	*	
38A3			oo	*	o	*	*	*		
38A5			oo	*	o	*	*	*		
39A1			o	*	o	o	*	*		*
39A3			*	*	*	*	*	*	◆	*
41A3			o	*	*	*	*	*		*
44A1			o	*	o	*	*	*		*
24A6			o							*
41A2			*						*	*
38A1			o							*
MAP adenomas analysed		10	27	24	24	24	24	24	6	21
MAP adenomas with mutations		n/a	21 (77.8%)	0	8 (33.3%)	2 (8.3%)	2 (8.3%)	1 (4.2%)	2 (33.3%)	5 (23.8%)

Number of adenomas analysed		20	69	62	62	62	62	62	23	47
Adenomas with mutations		n/a	49 (71%)	0	12 (19.4%)	5 (8.1%)	4 (6.5%)	2 (4.8%)	3 (13%)	5 (10.6%)

Table 4. Summary of Copy Number Variants (CNVs) detected by array CGH.

Adenoma	FAP/MAP	Location	CNV	Start	End	Size (bp)	OMIM Genes	HGNC Genes
24A3	MAP	15q11.1-15q21.1	GAIN	20.071.673	48.342.606	28.238.748	134	375
24A4		15q11.1-15q21.1	GAIN	20.071.673	48.342.606	28.238.748	134	375
37A1		8p23.1	DEL	6.805.940	9.615.505	2.809.566	15	74
		9q22.32	DEL	99.121.641	131.163.638	32.041.998	153	293
38A2		7p22.3 - 7q36.3	GAIN	54.215	157.723.016	157.668.802	589	1.243
44A2		8p23.1	DEL	7.691.931	8.046.302	354.372	3	15
		18p11.32	DEL	148.993	9.371.093	9.222.101	24	38
		18q21.1	DEL	47.594.529	78.012.800	30.418.272	70	104

Figure 1.



AACR American Association for Cancer Research

Clinical Cancer Research

Burden and profile of somatic mutation in duodenal adenomas from patients with familial adenomatous- and MUTYH-associated polyposis

Laura Elizabeth Thomas, Joanna J Hurley, Elena Meuser, et al.

Clin Cancer Res Published OnlineFirst August 8, 2017.